CAP6412
Advanced Computer Vision

Mubarak Shah
shah@crcv.ucf.edu
HEC-245
CAP6412, Spring 2023

• **Instructor**: Dr. Mubarak Shah
• **Email**: shah@crcv.ucf.edu
• **Office**: HEC 245
• **Phone**: 4078235077
• **Co-Instructor**: Dr. Tanvir Ahmed, Tanvir.Ahmed@ucf.edu
• **Time**: Mondays and Wednesdays 3:00 to 4:15PM
• **Location**: Zoom
• **Office Hours**: Mondays 2:00 to 3:00PM; Wednesdays 4:15 to 5:00PM; Fridays 1:00 to 2:00; and by appointment
• **Pre-requisite**: CAP5415
• **Course webpage**: [https://www.crcv.ucf.edu/courses/cap6412-spring-2023/](https://www.crcv.ucf.edu/courses/cap6412-spring-2023/)
Course Objective

• To expose graduate students to the cutting-edge research in Computer Vision.

• We will discuss research papers on diffusion models and their applications.
Students Learning Outcomes

• Read and understand a research paper.
• Write a comprehensive review of the paper.
• To identify strong and weak points of the paper.
• To generate own ideas to solve the same problem
• To work on research project and write a research paper
Grading Policy

Reports (individually, you have to do only 50% of the papers) 15%
Presentations (roughly two; by a group) 25%
Attendance (through zoom, please turn on your video, be on time) 10%
Project (by a group) 50%

Each group member will be evaluated by others; final grade on presentations and project will be scaled accordingly

No Mid Term, Final, Homework

• Late Policy
  • 0 for late reports
  • Projects
    • 20% off per day
    • up to 4 days
Project Timeline

- Three in-class presentations (see class schedule)
  - Project ideas / proposal [5%]
  - Update 1 [5%]
  - Update 2 [10%]
  - Final presentation/Paper [30%]
Programming/ GPU Cluster

- Pytorch, TensorFlow
- Each student will get an account on UCF HPC Newton
- Watch video: https://www.youtube.com/watch?v=niQ5hvABvKg&list=PLd3hlSJsX_km5il1HgmDB_z62BeoikFX&index=19
Papers/ Reports

• We will discuss one paper in each class
• All students will read the assigned paper before the class and write a report
• One group of students will make presentation and all students will participate in discussion
• Presentation should not be more than 30 minutes
• You can select the paper you want to present from the list on course webpage
• Reports will be due just before the class meeting through Web Courses

• Schedule Table: https://www.crcv.ucf.edu/courses/cap6412-spring-2023/schedule/
Reports (one page)

• Parts of Report
  • Summary
  • Good points
  • Weak points
  • Questions
  • Ideas

• Reports will not be graded

• We will follow the honor system, by submitting report you will pledge that you have read the paper, have written report yourself, have not copied word by word from the paper and from any other student or resources like ChatGPT etc
Statement of Academic Integrity

- The UCF Golden Rule (http://goldenrule.sdes.ucf.edu/) will be observed in the class. Plagiarism and Cheating of any kind on an examination, quiz, or assignment will result at least in an "F" for that assignment (and may, depending on the severity of the case, lead to an "F" for the entire course) and may be subject to appropriate referral to the Office of Student Conduct for further action. I will assume for this course that you will adhere to the academic creed of this University and will maintain the highest standards of academic integrity. In other words, don't cheat by giving answers to others or taking them from anyone else. I will also adhere to the highest standards of academic integrity, so please do not ask me to change (or expect me to change) your grade illegitimately or to bend or break rules for one person that will not apply to everyone.
Paper presentation review/Rehearsal Schedule

• For Monday presentation
  • Slide Review: Wednesday 4:15 a week before the scheduled presentation
  • Rehearsal: Friday a week before the scheduled presentation 1:00PM during Office hours

• For Wednesday presentation
  • Slide Review: A week before the scheduled presentation: Friday 1:00PM during Office hours
  • Rehearsal: A week of presentation on Monday 2:00PM during Office hours
There is no textbook for this class. We will discuss recent research papers.

A good deep learning textbook:

Other Resources

• Survey paper
  
• https://yang-song.net/blog/2021/score/
• https://www.youtube.com/watch?v=cS6JQpEY9cs
Research Paper

• Each paper has following parts
  1. Title
  2. Abstract
  3. Introduction
  4. Rest of the paper
     • Related Work
     • Method
     • Results
     • Conclusion

• Each part is equally important (25% each!)
How to read a research paper?

- You must read the paper several times to understand it.
  - When you read the paper first time, if you do not understand something do not get stuck, keep reading assuming you will figure out that later.
  - When you read it the second time, you will understand much more, and the third time even more ...
- Read the abstract first then look at the figures with captions and then conclusion
How to read a research paper?

• Try first to get a general idea of the paper
  • What problem is being solved?
  • What are the main steps?
  • How can I implement the method?, Even though I do not understand why each step is performed the way it is performed

• Try to relate the method to other methods you know, and conceptually find similarities and differences.
How to read a research paper?

• In the first reading it may be a good idea to skip the related work.

• Do not use dictionary to just look up the meaning of technical terms.

• Try to understand each concept in isolation, and then integrate them to understand the whole paper.
Useful Blogs about how to read a paper?

• [https://web.stanford.edu/class/ee384m/Handouts/HowtoReadPaper.pdf](https://web.stanford.edu/class/ee384m/Handouts/HowtoReadPaper.pdf)

• [https://blogs.lse.ac.uk/impactofsocialsciences/2016/05/09/how-to-read-and-understand-a-scientific-paper-a-guide-for-non-scientists/](https://blogs.lse.ac.uk/impactofsocialsciences/2016/05/09/how-to-read-and-understand-a-scientific-paper-a-guide-for-non-scientists/)
Short Introduction

• Thanks to
  • Navid Kardan, Ming Li, ...

• Material taken from different resources including:
GPT (Generative Pre-trained Transformer)

• Language model that uses deep learning to produce human-like text.

• 2020: GPT-3
  • 175 billion parameters
  • capable of machine translation, text generation, semantic analysis

• 2021: DALL-E
  • Text to image

• 2021: CLIP (Contrastive Language Image Pre-training)

• 2021: GLIDE (Diffusion Model)

• 2022: DALLE-2 (UNCLIP) (Diffusion Model)

• 2022: Chat GPT
  • interacts in a conversational way

• GPT4:?
Text-to-Image Generation: Timeline

EBMs
Flow-based models
GANs
VAEs

GLIDE
by OpenAI
December 2021

DALL-E
by OpenAI
January 2021

DALL-E 2
by OpenAI
April 2022

Imagen
by Google Brain
May 2022

Short timeline of image generation and text-to-image [1]
Generative Models

Diffusion Models

Tackling the Generative Learning Trilemma with Denoising Diffusion GANs, ICLR 2022
Diffusion Applications

- Image generation
- Text-to-image
- Super resolution
- Image editing
- Inpainting
- Image segmentation
- Video Generation
- .....
Unconditional Image Generation

Conditional Image Generation

Text-to-Image

A small gray bird with white and dark gray wingbars and white breast
This beautiful little bird has a white breast and very intriguing red eyes
A small sized blue bird that has a short pointed bill
The long wings spreaded showing the breast and the belly of the large bird
An airplane that is parked at airport
A giraffe is standing in a green field
Some children are playing soccer on the field
A white and blue bus driving down a road next to trees

Super-resolution

Segmentation

Video Generation

Inpainting

References

1024 x 1024 samples generated from score-based models [20]
Samples from the NCSNv2 [18] model. From left to right: FFHQ 256x256, LSUN bedroom 128x128, LSUN tower 128x128, LSUN church, outdoor 96x96, and CelebA 64x64.
1024 x 1024 samples from a score-based model trained on the FFHQ dataset.

Some additional (uncurated) samples for other datasets (taken from this GitHub repo):
256 x 256 samples on LSUN bedroom.
Inpainting
Image Colorization
Image Colorizations

We can even colorize gray-scale portraits of famous people in history (Abraham Lincoln) with a time-dependent score-based model trained on FFHQ. The image resolution is 1024 x 1024.
Stable Diffusion

High-Resolution Image Synthesis with Latent Diffusion Models

Robin Rombach\(^1\) * Andreas Blattmann\(^1\) * Dominik Lorenz\(^1\) Patrick Esser\(^\circ\) Björn Ommer\(^1\)

\(^1\)Ludwig Maximilian University of Munich & IWR, Heidelberg University, Germany \(^\circ\)Runway ML

https://github.com/CompVis/latent-diffusion
Text-to-Image

'A watercolor painting of a chair that looks like an octopus'

'A shirt with the inscription：“I love generative models!”'
DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation

Nataniel Ruiz\textsuperscript{1,2}, Yuanzhen Li\textsuperscript{1}, Varun Jampani\textsuperscript{1}, Yael Pritch\textsuperscript{1}, Michael Rubinstein\textsuperscript{1}, and Kfir Aberman\textsuperscript{1}

\textsuperscript{1}Google Research \quad \textsuperscript{2}Boston University
Motivation: Generate new context for single instance

- in the Acropolis
- in a doghouse
- in a bucket
- getting a haircut

- worn by a bear
- at Mt. Fuji
- on top of snow
- with Eiffel Tower
Expression modification (“A [state] [V] dog”)

depressed  sleeping  sad  joyous
barking  crying  frowning  screaming
Color modification ("A [color] [V] car")

Hybrids ("A cross of a [V] dog and a [target species]")
Person Image Synthesis via Denoising Diffusion Model

Ankan Kumar Bhunia\textsuperscript{1}  Salman Khan\textsuperscript{1,2}  Hisham Cholakkal\textsuperscript{1}  Rao Muhammad Anwer\textsuperscript{1,4}  
Jorma Laaksonen\textsuperscript{4}  Mubarak Shah\textsuperscript{5}  Fahad Shahbaz Khan\textsuperscript{1,3}  
\textsuperscript{1}Mohamed bin Zayed University of AI, UAE  \textsuperscript{2}Australian National University, Australia  
\textsuperscript{3}Linköping University, Sweden  \textsuperscript{4}Aalto University, Finland  \textsuperscript{5}University of Central Florida, USA
DiffusionDet: Diffusion Model for Object Detection

Shoufa Chen\textsuperscript{1}, Peize Sun\textsuperscript{1}, Yibing Song\textsuperscript{2}, Ping Luo\textsuperscript{1}
\textsuperscript{1}The University of Hong Kong  \textsuperscript{2}Tencent AI Lab
\{sfchen, pzsun, pluo\}@cs.hku.hk  yibingsong.cv@gmail.com
SOLVING INVERSE PROBLEMS IN MEDICAL IMAGING WITH SCORE-BASED GENERATIVE MODELS

Yang Song*, Liyue Shen*, Lei Xing & Stefano Ermon
Stanford University
{yangsong@cs,liyues@,lei@,ermon@cs}.stanford.edu
Improved generation

Medical image reconstruction

- X-ray source
- Gantry
- Patient
- Table
- Multiple row detector
Medical image reconstruction

- X-ray source
- Gantry
- Patient
- Table
- Multiple row detector

Sparse-view sinogram
Medical image reconstruction

Forward model $p(y | x)$ is given by physical simulation

Sparse-view computed tomography (CT)

Cross-sectional image
Medical image reconstruction

Sparse-view CT (just 23 projections)

PSNR: 20.30, SSIM: 0.778  PSNR: 22.78, SSIM: 0.603  PSNR: 31.76, SSIM: 0.882  PSNR: 35.23, SSIM: 0.912

FISTA  Neumann network  SIN-PRN  Ours  Ground truth

Outperforms deep learning methods specifically trained for 23 projections

[Song et al. ICLR 2022]
Introductions

• My Introduction

• Your Introduction
Thankyou